The PubChem Compound Database

A catalog of chemical compounds from various sources

https://www.ncbi.nlm.nih.gov/pccompound & https://pubchem.ncbi.nlm.nih.gov/

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope and Access

The PubChem Compound database is a standardized, consolidated set of chemical compound records. These records contain biological and physicochemical properties information as well as links to other NCBI and external databases. Links to several visualization and analysis tools are also provided. PubChem Compound records can be searched with text queries by NCBI's Entrez Search Engine and with chemical structural information by the PubChem Chemical Structure Similarity Search Tool. Several venues are available for accessing data from the PubChem family of databases.

Web interfaces:

PubChem Homepage	https://pubchem.ncbi.nlm.nih.gov (**A**)
PubChem Compound Database Homepage	https://www.ncbi.nlm.nih.gov/pccompound (**B**)

FTP or the bulk download tool:

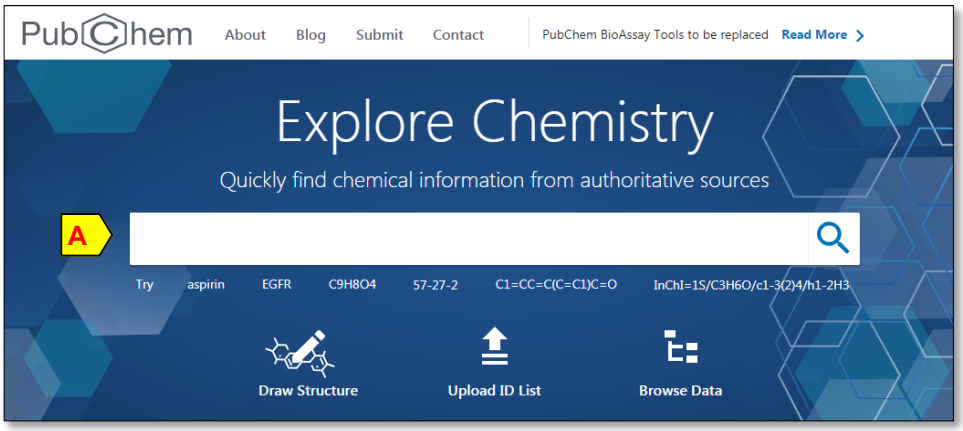
PubChem FTP	https://ftp.ncbi.nlm.nih.gov/pubchem/
PC Structure download	https://pubchem.ncbi.nlm.nih.gov/pc_fetch/pc_fetch.cgi

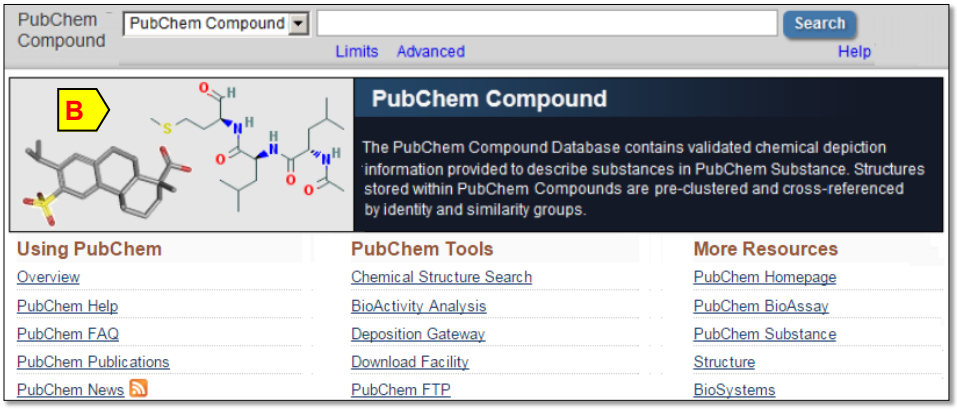
Programmatic approaches:

Entrez Program Utilities	https://www.ncbi.nlm.nih.gov/books/NBK25501/
Power User Gateway and REST	https://pubchemdocs.ncbi.nlm.nih.gov/programmatic-access

The PubChem Project

The PubChem Project was originally developed in 2004 to serve as a freely accessible repository and point of data access for the NIH Molecular Libraries Roadmap Initiative. It was designed for use as a research tool to assist in the development of new medications. As part of the powerful family of databases operated by NCBI, including GenBank, Gene, BioSystems and PubMed, the integration of these databases makes the whole much greater than the sum of its parts. The database connects chemical information with biomedical research and clinical information, organizing facts from and links to numerous databases in a unified system.



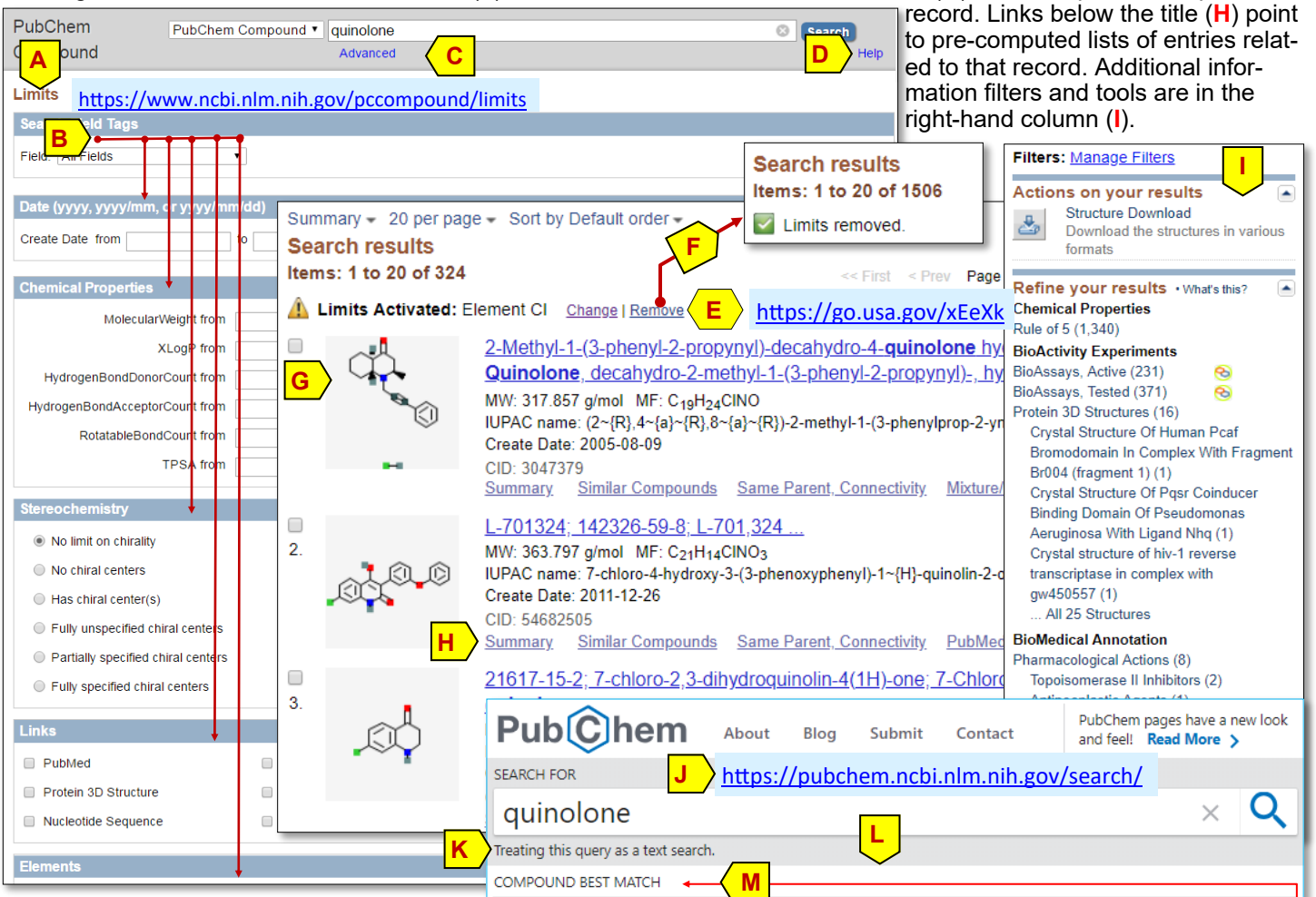


The PubChem system of three interlinked databases (PC BioAssay, PC Substance, and PC Compound) is fully integrated into NCBI's Entrez system, making database records fully indexed and searchable using standard Entrez search methods. Data in PubChem has been submitted to NCBI by researchers, external databases, and vendors and is stored within the primary databases, PC BioAssay and PC Substance. The PubChem group performs several key functions in the production of compound records: it standardizes, consolidates, and calculates physicochemical properties; provides additional descriptive information; and computes, clusters and cross-references other structurally similar compounds. Thus, PC Compound records serve as a derivative reference database containing validated chemical information with links to corresponding PC Substance and PC BioAssay records, as well as to other NCBI and external databases which provide additional information on biological and physicochemical properties and potential biological effects and mechanisms of action.

PubChem also provides several visualization and analysis tools including the Chemical Structure Search algorithm and Structure Clustering tool, 2D structure and 3D conformer viewers, interactive tabular and graphical plotting interfaces for examining BioActivity, and a Structure-Activity Heatmap which allows for interactive visualization of data produced from both Structural and BioAssay Clustering.

Finding Chemical Information in PubChem

Searching with terms from the PC Compound homepage retrieves relevant compounds matching the input query. The Limits page (**A**), under the "Advanced Search" link from the PubChem homepage, allows access to commonly used criteria (**B**) for more focused searches. The Advanced page (**C**, details not shown) provides access to indexed fields and index terms. The help documentation (**D**) contains a detailed list of indices of the database. The search results is obtained from searching with "quinolone" limited to chlorine-containing compounds as indicated by the message (**E**). Removing these limits retrieves more records (**F**). The structure thumbnails and the titles (**G**) link to a specific compound record. Links below the title (**H**) point to pre-computed lists of entries related to that record. Additional information filters and tools are in the right-hand column (**I**).



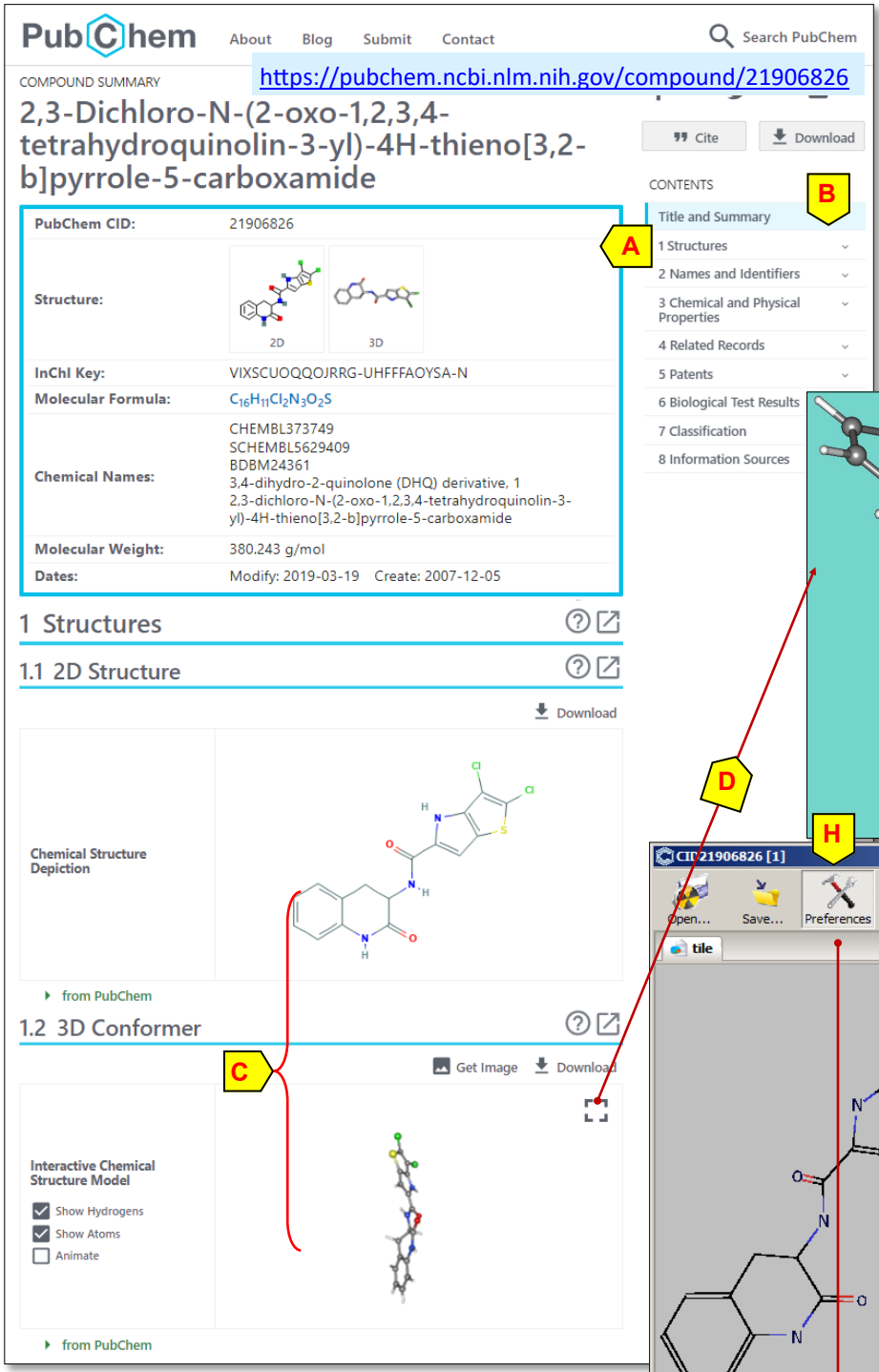
The new search interface

The newly introduced search interface (**J**) added additional functionality to streamline the search, chief among them is the automatic selection of search algorithm to better match the input query. The system alert users about the decision, as shown by the same text query with "quinolone" (**K**).

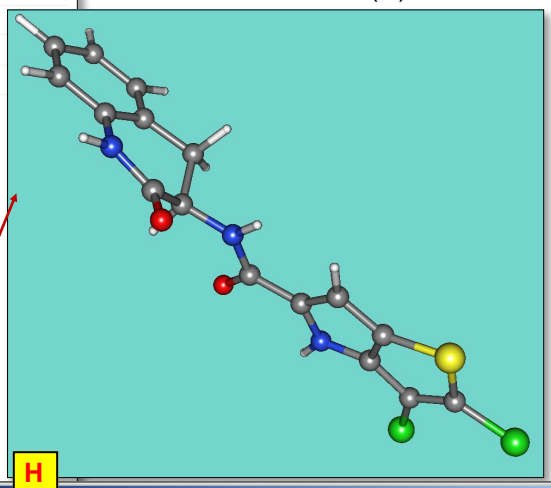
This interface also presents search results more intuitively (**L**) so records retrieved are better grouped and more readable by

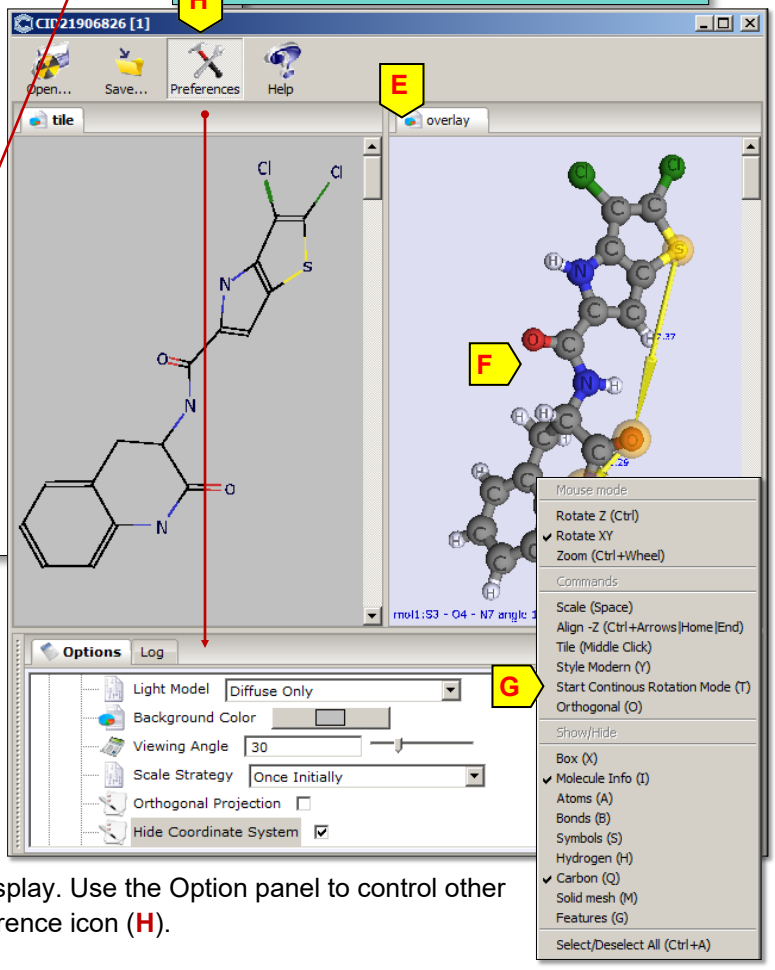
- placing the best matched record (**M**) at the very top if such record is found
- listing related records within PubChem and other databases under easy to reach tabs (**N**)
- making the filter and sorting functions (**O**) readily accessible to allow further refining of the result display, and
- providing additional links in the right column (**P**)

Displaying PubChem Compound Records



The PubChem Compound record display adopts a book-like format that organizes the content into numbered sections. The top section (**A**) provides a brief summary of the record, which includes a list of synonyms. A collapsible table of contents (**B**) helps navigation among different sections, click a section title to jump to that section. Graphical presentation of the structure in 2D and 3D formats are shown in Sections 1 (**C**). Click the "Download" link to download the structure file. Click the expansion icon to interactively examine the 3D structure in full screen mode (**D**) .



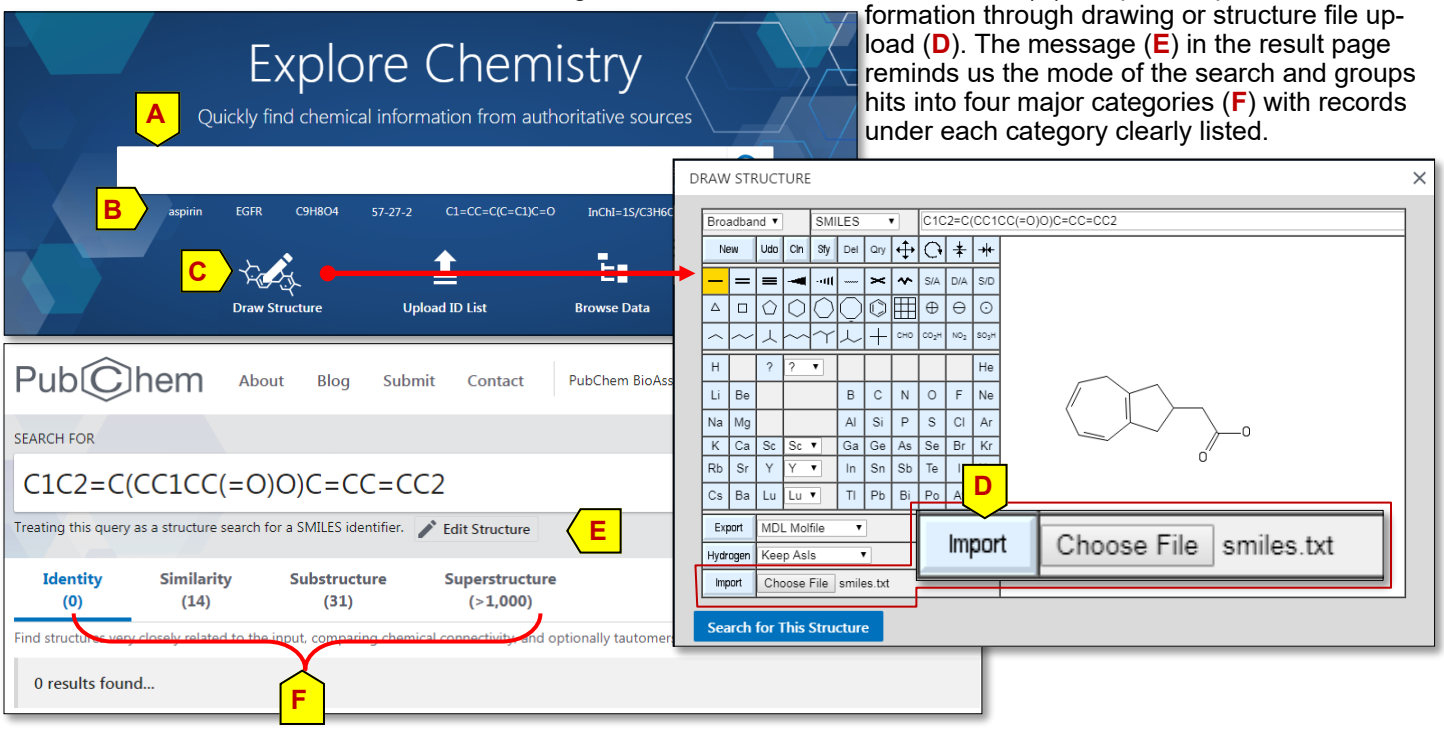


PubChem Compound 3D Viewing Tools

You can interact with saved 3D structure using the pc3d tool (https://pubchem.ncbi.nlm.nih.gov/pc3d/). There you can use the tile and overlay panels (**E**) to examine the structure from different perspectives. The overlay panel allows more detailed examination, such as measuring the distance between two selected atoms (**F**). Right click to see a menu (**G**) for customization of the display. Use the Option panel to control other general settings, which is activated by clicking the Preference icon (**H**).

Searching with Structures to Find Chemical Information

The PubChem search box (**A**) takes input in various formats, with sample input linked below (**B**). It supports text query as well as structure search. You can use drawing tool under the "Draw Structure" icon (**C**) to input complex structure information through drawing or structure file upload (**D**). The message (**E**) in the result page reminds us the mode of the search and groups hits into four major categories (**F**) with records under each category clearly listed.



Programmatic Data Access

PubChem supports programmatic access to the data it hosts. The table below sums up available services along with links to detailed help document for more information. Each of the services has its own set of functions and strength. You should select the services that strikes a balance among various factors, such as specific needs, capability on hand, and the nature of the project.

Service Name	Functions
Entrez Programming Utilities	Entrez styled text search Summary information Links to relevant records in other databases More info: https://www.ncbi.nlm.nih.gov/books/NBK25501
PUG	Data exchange service, operating through XML and HTTP POST Help manual: https://pubchemdocs.ncbi.nlm.nih.gov/power-user-gateway
PUG-SOAP	A simple object access protocol-based services for SOAP-aware applications More info: https://pubchemdocs.ncbi.nlm.nih.gov/pug-soap
PUG-REST	A Representational State Transfer (REST)-style web service access layer to PubChem, provides a simplified access route to PubChem without the overhead of XML or SOAP envelope More info: https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest
PUG-VIEW	A REST-style web service that provides access to annotation information for PubChem Substance, Compound, or BioAssay records. More info: https://pubchemdocs.ncbi.nlm.nih.gov/pug-view
RDF	The PubChemRDF project provides Resource Description Framework (RDF) formatted information for the PubChem Compound, Substance, and Bioassay databases. More info: https://pubchemdocs.ncbi.nlm.nih.gov/rdf

Help Documents

An extensive set of help document for PubChem is online at https://pubchemdocs.ncbi.nlm.nih.gov/about